

PA12.1 – Variability in idiomatic multiword expressions

Student: Irene Pagliai Supervisors: Anke Holler, Caroline Sporleder

Ext./Th.Com.: Thomas Weskott, Alessandro Lenci (Pisa)

I. The form-meaning mismatch

- Idioms: multiword expressions (MWEs) with a conventional figurative meaning. [1]
 - essere alla frutta* (lit. “be at the fruit”) = “be at the end of one’s resources”
- Not a homogeneous class! [2] Idioms vary across dimensions such as *literal plausibility*, *decomposability* and *transparency* → **intra-idiom variation**.
- Idioms may also appear in contexts where their literal meaning is pivotal [3]: literal-compositional contexts and idiomatic wordplay → **extra-idiom variation**.

Overall research aim

- To investigate the complexity of idioms examining them both as *types* (inherent characteristics) and *tokens* (contextual instantiations).

Research questions

Type-based perspective:

- Are idiom-internal feature correlations cross-linguistically replicable? How do lexical/syntactic similarity affect rating variation?

Token-based perspective:

- How do idiom features influence integration in literal and wordplay contexts?

II. Methodology and hypotheses

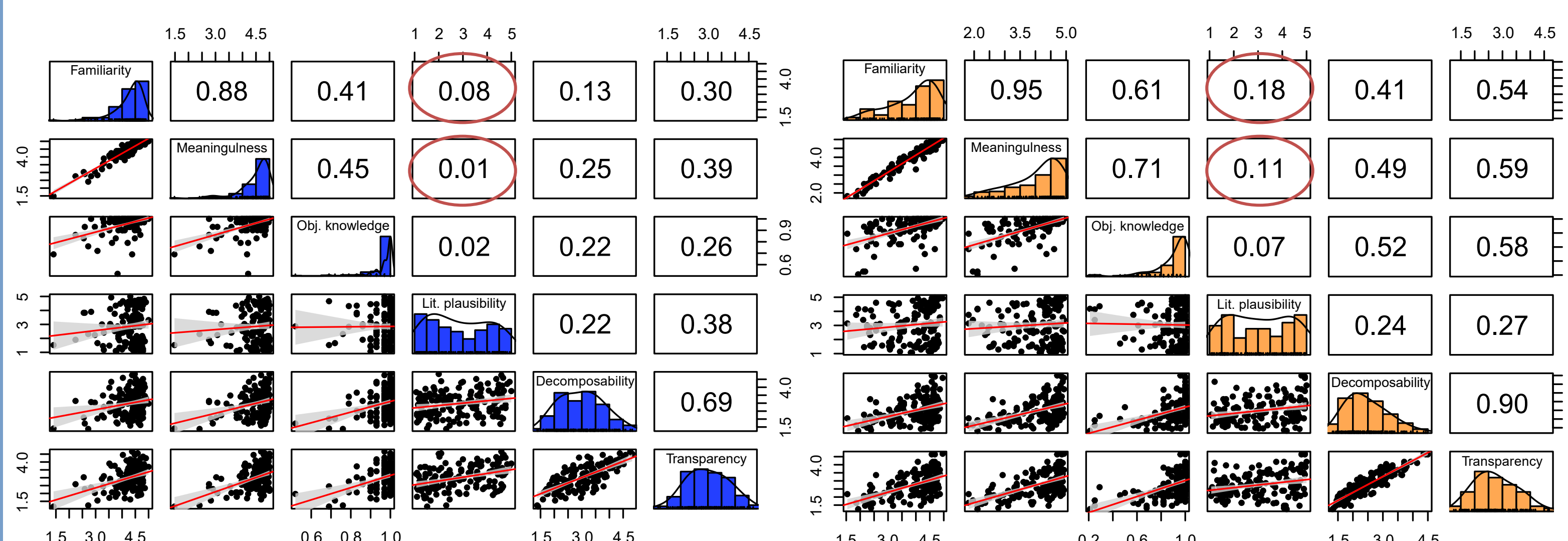
- Dataset:** normed lexicon of 150 Italian–English idiom pairs with comparable meanings, varying in lexical overlap (LL, SL, PL [4]) and syntactic structure (shallow parsing).
- Norming:** [5] speakers rated each idiom for:
 - Experience-Based Variables (EBVs): *familiarity*, *meaningfulness*, *objective knowledge*
 - Content-Based Variables (CBVs): *literal plausibility*, *decomposability*, *transparency* → combined to derive the *Potential Idiomatic Ambiguity* (PIA) index.
- Experiment:** acceptability-rating task on 16 Italian idioms (high/low PIA) embedded in fictional dialogues. Two conditions: literal-compositional contexts & figurative-wordplay contexts. Participants rated how well idiom literal completions integrated into each context; reading times were also recorded.

Hypotheses:

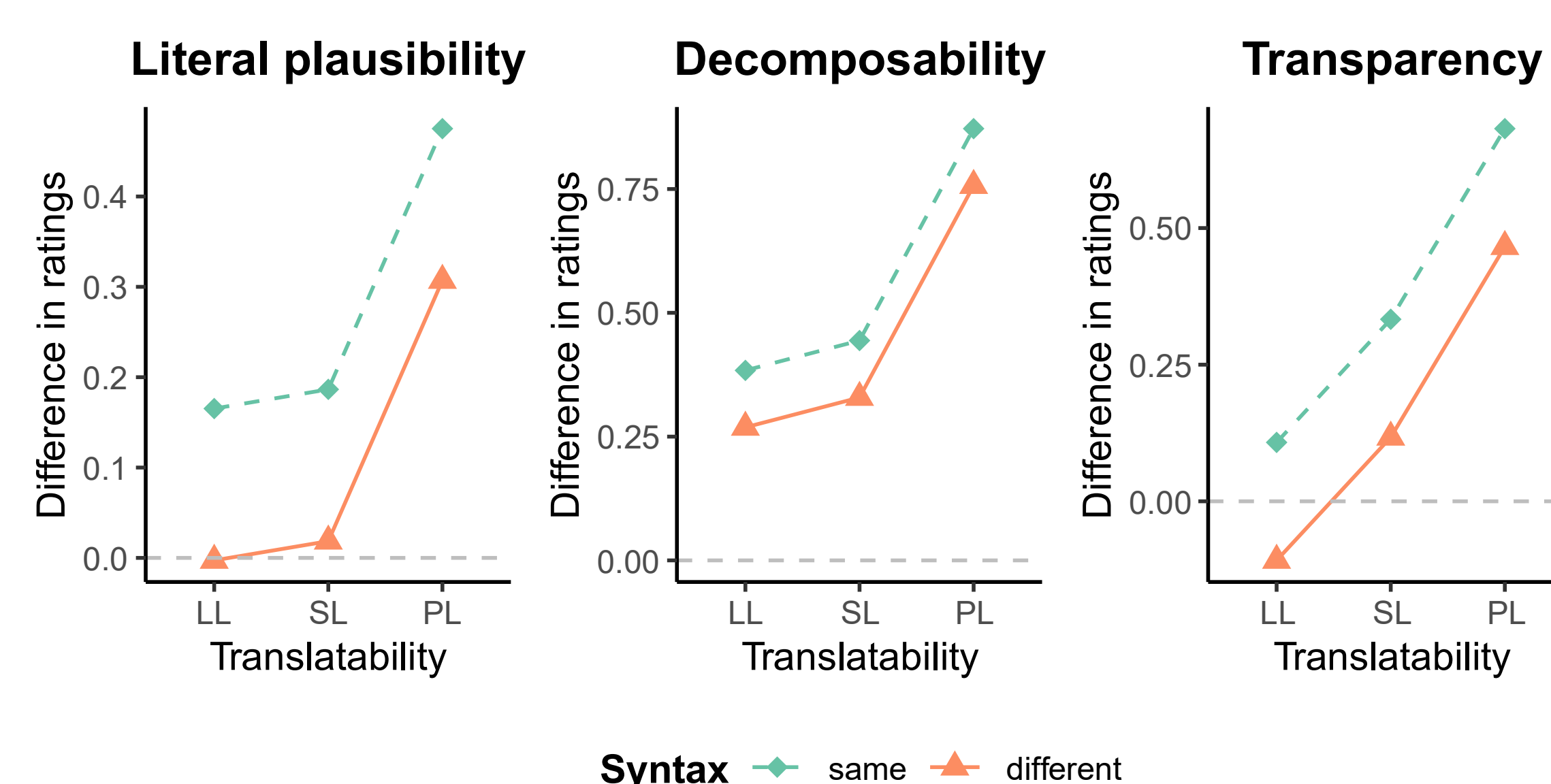
- Type perspective:** (i) correlational patterns expected to replicate across languages; (ii) CBV ratings predicted to diverge more from LL→SL→PL; syntactic match expected to impact decomposability only.
- Token perspective:** (iii) High PIA → smoother integration in literally-relevant contexts (compositional & wordplay).

III. Results and discussion

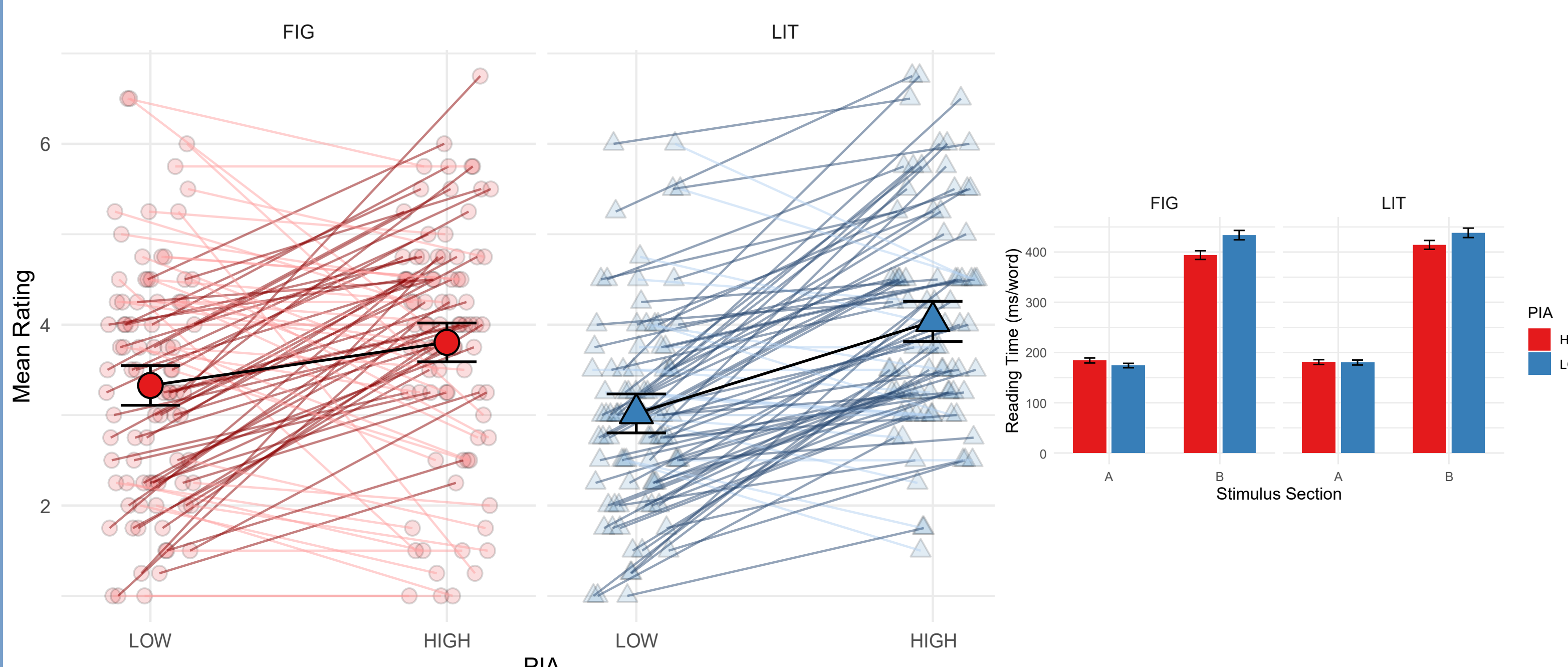
- Correlations are largely replicable across **Italian** and **English**. Interesting divergence in literal plausibility–EBVs.



- CBV rating difference increases from LL→SL→PL. Syntactic mismatch reduces differences, but it’s never significant.



- High PIA → higher mean acceptability & lower reading times of literal completions in both contexts, but greater variability across participants in **wordplay** contexts!



IV. Consequences and follow-up questions

- Type perspective:** (i) idioms with comparable meanings show consistently structured representations in speakers’ mental lexicons. (ii) The cross-linguistic similarity gradient is lexically driven (test finer-grained syntactic annotation).
- Token perspective:** (iii) idiom-internal features modulate contextual integration → positive effect of PIA. Higher variability in wordplay warrants more items and focus on inter-individual divergences in interpretive strategies.
 - **Project 12.2** extends the analysis of conventionalized linguistic items to classifier constructions in sign languages and co-speech gestures → how conventional handshapes combine with iconic and context-dependent features.